7874

REGISTRY OF PATENTS
SINGAPORE

This is to certify that the annexed is a true copy of following application as filed with the Registry.

Date of Filing : 7 MARCH 2003

Application Number : 200301300-0

Applicant(s) /
Proprietor(s) of Patent : STMICROELECTRONICS ASIA PACIFIC PTE LTD

Title of Invention : DEVICE AND PROCESS FOR USE IN ENCODING AUDIO DATA

CERTIFIED COPY OF PRIORITY DOCUMENT

SHARMAINE WU (Ms)
Assistant Registrar
for REGISTRAR OF PATENTS
SINGAPORE

24 MARCH 2004

•G00001•

INTELLECTUAL PROPERTY OFFICE OF SINGAPORE

REQUEST FOR THE GRANT OF A PATENT UNDER SECTION 25

101101

---

* denotes mandatory fields

---

1. YOUR REFERENCE* | 1013759PAT/STMicro/MK/FL

---

2. TITLE OF INVENTION* | DEVICE AND PROCESS FOR USE IN ENCODING AUDIO DATA

---

3. DETAILS OF APPLICANT(S)* (see note 3)   Number of applicant(s) | 1 |

(A) Name | STMicroelectronics Asia Pacific Pte Ltd
(A Private Limited Company Incorporated In The Republic Of Singapore)

Address | 28 ANG MO KIO INDUSTRIAL PARK 2, SINGAPORE 569508

State | | Country | Singapore

| X | For corporate applicant | | For individual applicant

State of incorporation | | State of residency | |

Country of incorporation | Singapore | Country of residency | |

| | For others (please specify in the box provided below)

(B) Name | |

Address | |

State | | Country | |

•ACTION•

☐ For corporate applicant   ☐ For individual applicant

State of incorporation [   ]   State of residency [   ]

Country of incorporation [   ]   Country of residency [   ]

☐ For others (please specify in the box provided below)

[   ]

(C) Name   [   ]

Address   [   ]

State   [   ]   Country [   ]

☐ For corporate applicant   ☐ For individual applicant

State of incorporation [   ]   State of residency [   ]

Country of incorporation [   ]   Country of residency [   ]

☐ For others (please specify in the box provided below)

[   ]

☐   Further applicants are to be indicated on continuation sheet 1

---

4. DECLARATION OF PRIORITY (see note 5)

A. Country/country designated [   ]

                                                                DD MM YYYY
File number [   ]   Filing Date [   ]

B. Country/country designated [   ]

                                                                DD MM YYYY
File number [   ]   Filing Date [   ]

☐   Further details are to be indicated on continuation sheet 6

---

5. INVENTOR(S)* (see note 6)

A The applicant(s) is/are the sole/joint inventor(s)   Yes ☐   No ☒

B. A statement on Patents Form 8 is/will be furnished   Yes ☐   No ☒

6. CLAIMING AN EARLIER FILING DATE UNDER (see note 7)

☐ section 20(3)   ☐ section 26(6)   ☐ section 47(4)

Patent application number [   ]

DD MM YYYY

Filing Date [   ]

Please mark with a cross in the relevant checkbox provided below
(Note: Only one checkbox may be crossed.)

☐ Proceedings under rule 27(1)(a)

DD MM YYYY

Date on which the earlier application was amended [   ]

☐ Proceedings under rule 27(1)(b)

7. SECTION 14(4)(C) REQUIREMENTS (see note 8)

Invention has been displayed at an international exhibition.   Yes ☐   No ☒

8. SECTION 114 REQUIREMENTS (see note 9)

The invention relates to and/or used a micro-organism deposited for the purposes of disclosure in accordance with section 114 with a depository authority under the Budapest Treaty.

Yes ☐   No ☒

9. CHECKLIST*

(A) The application consists of the following number of sheets

i.	Request	5	Sheets
ii	Description	14	Sheets
iii.	Claim(s)	4	Sheets
iv.	Drawing(s)	3	Sheets
v.	Abstract (Note. The figure of the drawing, if any, should accompany the abstract)	1	Sheets
Total number of sheets		27	Sheets

(B) The application as filed is accompanied by.

☐ Priority document(s)   ☐ Translation of priority document(s)

☐ Statement of Inventorship & right to grant ☐ International exhibition certificate

10. DETAILS OF AGENT (see notes 10, 11 and 12)

Name	
Firm	DONALDSON & BURKINSHAW

11. ADDRESS FOR SERVICE IN SINGAPORE* (see note 10)

Block/Hse No. [ ]   Level No. [ ]   Unit No./PO Box [3667]

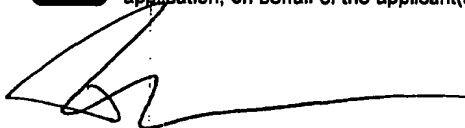Street Name [ ]

Building Name [ ]

Postal Code [905667]

12. NAME, SIGNATURE AND DECLARATION (WHERE APPROPRIATE) OF APPLICANT OR AGENT* (see note 12)
(Note: Please cross the box below where appropriate.)

[X] I, the undersigned, do hereby declare that I have been duly authorised to act as representative, for the purposes of this application, on behalf of the applicant(s) named in paragraph 3 herein.

_____
Michael S. Kraal
Name and Signature

DD MM YYYY
[06-03-2003]

NOTES:

1. This form when completed, should be brought or sent to the Registry of Patents together with the rest of the application. Please note that the filing fee should be furnished within the period prescribed.

2. The relevant checkboxes as indicated in bold should be marked with a cross where applicable.

3. Enter the name and address of each applicant in the spaces provided in paragraph 3.
    Where the applicant is an individual
    - Names of individuals should be indicated in full and the surname or family name should be underlined.
    -The address of each individual should also be furnished in the space provided.
    - The checkbox for "For individual applicant" should be marked with a cross

    Where the applicant is a body corporate
    - Bodies corporate should be designated by their corporate name and country of incorporation and, where appropriate, the state of incorporation within that country should be entered where provided.
    - The address of the body corporate should also be furnished in the space provided.
    - The checkbox for "For corporate applicant" should be marked with a cross

    Where the applicant is a partnership
    - The details of all partners must be provided. The name of each partner should be indicated in full and the surname or family name should be underlined.
    - The address of each partner should also be furnished in the space provided.
    - The checkbox for "For others" should be marked with a cross and the name and address of the partnership should be indicated in the box provided.

4. In the field for "Country", please refer to the standard list of country codes made available by the Registry of Patents and enter the country code corresponding to the country in question.

5. The declaration of priority in paragraph 4 should state the date of the previous filing, the country in which it was made, and indicate the file number, if available. Where the application relied upon in an International Application or a regional patent application e g European patent application, one of the countries designated in that application [being one falling under section 17 of the Patents Act] should be identified and the country should be entered in the space provided

6. Where the applicant or applicants is/are the sole inventor or the joint inventors, paragraph 5 should be completed by marking with a cross the 'YES' checkbox in the declaration (A) and the 'NO' checkbox in the alternative statement (B) Where this is not the case, the 'NO' checkbox in declaration (A) should be marked with a cross and a statement will be required to be filed on Patents Form 8.

7. When an application is made by virtue of section 20(3), 26(6) or 47(4), the appropriate section should be identified in paragraph 6 and the number of the earlier application or any patent granted thereon identified. Applicants proceeding under section 26(6) should identify which provision in rule 27 they are proceeding under. If the applicants are proceeding under rule 27(1)(a), they should also indicate the date on which the earlier application was amended.

8   Where the applicant wishes an earlier disclosure of the invention by him at an International Exhibition to be disregarded in accordance with section 14(4)(c), then the 'YES' checkbox at paragraph 7 should be marked with a cross. Otherwise, the 'NO' checkbox should be marked with a cross

9. Where in disclosing the invention the application refers to one or more micro-organisms deposited with a depository authority under the Budapest Treaty, then the 'YES' checkbox at paragraph 8 should be marked with a cross. Otherwise, the 'NO' checkbox should be marked with a cross. Attention is also drawn to the Fourth Schedule of the Patents Rules.

10. Where an agent is appointed, the fields for "DETAILS OF AGENT" and "ADDRESS FOR SERVICE IN SINGAPORE" should be completed and they should be the same as those found in the corresponding Patents Form 41 In the event where no agent is appointed, the field for "ADDRESS FOR SERVICE IN SINGAPORE" should be completed, leaving the field for "DETAILS OF AGENT" blank.

11. In the event where an individual is appointed as an agent, the sub-field "Name" under "DETAILS OF AGENT" must be completed by entering the full name of the individual. The sub-field "Firm" may be left blank. In the event where a partnership/body corporate is appointed as an agent, the sub-field "Firm" under "DETAILS OF AGENT" must be completed by entering the name of the partnership/body corporate. The sub-field "Name" may be left blank.

12. Attention is drawn to sections 104 and 105 of the Patents Act, rules 90 and 105 of the Patents Rules, and the Patents (Patent Agents) Rules 2001.

13. Applicants resident in Singapore are reminded that if the Registry of Patents considers that an application contains information the publication of which might be prejudicial to the defence of Singapore or the safety of the public, it may prohibit or restrict its publication or communication. Any person resident in Singapore and wishing to apply for patent protection in other countries must first obtain permission from the Singapore Registry of Patents unless they have already applied for a patent for the same invention in Singapore. In the latter case, no application should be made overseas until at least 2 months after the application has been filed in Singapore, and unless no directions had been issued under section 33 by the Registrar or such directions have been revoked. Attention is drawn to sections 33 and 34 of the Patents Act.

14. If the space provided in the patents form is not enough, the additional information should be entered in the relevant continuation sheet. Please note that the continuation sheets need not be filed with the Registry of Patents if they are not used.

DEVICE AND PROCESS FOR USE IN ENCODING AUDIO DATA

FIELD OF THE INVENTION

The present invention relates to a device and process for use in encoding audio data, and in
5   particular to a psychoacoustic mask generation process for MPEG audio encoding.

BACKGROUND

The MPEG-1 audio standard, as described in the International Standards Organisation
(ISO) document ISO/IEC 11172-3: Information technology - *Coding of moving pictures
10 and associated audio for digital storage media at up to about 1.5Mbps* ("the MPEG-1
standard"), defines processes for lossy compression of digital audio and video data. The
MPEG-1 standard defines three alternative processes or "layers" for audio compression,
providing progressively higher degrees of compression at the expense of increasing
complexity. The second layer, referred to as MPEG-1-L2, provides an audio compression
15 format widely used in consumer multimedia applications. As these applications progress
from providing playback only to also providing recording, a need arises for consumer-
grade and consumer-priced devices that can generate MPEG-1-L2 compliant audio data.

The reference implementation for an MPEG-1-L2 encoder described in the MPEG-1
20 standard is not suitable for real-time consumer applications, and requires considerable
resources in terms of both memory and processing power.   In particular, the
psychoacoustic masking process used in the MPEG-1-L2 audio encoder referred to uses a
number of successive and processing intensive power and energy data conversions that
also incur a repeated loss in precision.
25

Accordingly, it is desired to address the above or at least provide a useful alternative.

SUMMARY OF THE INVENTION

In accordance with the present invention there is provided a mask generation process for use in encoding audio data, including:

generating linear masking components from said audio data;

5   generating logarithmic masking components from said linear masking components; and

generating a global masking threshold from the logarithmic masking components.

The present invention also provides a mask generation process for use in encoding audio data, including:

10   generating respective masking thresholds from logarithmic masking components using a masking function of the form:

$$vf = -17 * dz, \quad 0 \le dz < 8$$

The present invention also provides a mask generation process for use in encoding audio

15   data, including:

generating a global masking threshold from logarithmic masking components according to:

$$LT_g(i) = \max\left[LT_q(i) + \max_{j=1}^{m}\left\{LT_{tonal}\left[z(j), z(i)\right]\right\} + \max_{j=1}^{n}\left\{LT_{noise}\left[z(j), z(i)\right]\right\}\right]$$

where i and j are indices of spectral audio data, $z(i)$ is a Bark scale value for spectral

20   line i, $LT_{tonal}\left[z(j), z(i)\right]$ is a tonal masking threshold for lines i and j, $LT_{noise}\left[z(j), z(i)\right]$ is a non-tonal masking threshold for lines i and j, m is the number of tonal spectral lines, and n is the number of non-tonal spectral lines.

The present invention also provides a mask generator for an audio encoder, said mask

25   generator adapted to generate linear masking components from input audio data, logarithmic masking components from said linear masking components; and a global masking threshold from the logarithmic masking components.

The present invention also provides a psychoacoustic masking process for use in an audio encoder, including:

generating energy values from Fourier transformed audio data;

determining sound pressure level values from said energy values;

5 selecting tonal and non-tonal masking components on the basis of said energy values;

generating power values from said energy values;

generating masking thresholds on the basis of said masking components and said power values; and

generating signal to mask ratios for a quantizer on the basis of said sound pressure

10 level values and said masking thresholds.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the present invention are hereinafter described, by way of example only, with reference to the accompanying drawings, wherein:

15 Figure 1 is a block diagram of a preferred embodiment of an audio encoder;

Figure 2 is a flow diagram of a prior art process for generating masking data;

Figure 3 is a flow diagram of a mask generation process executed by a mask generator of the audio encoder.

20 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As shown in Figure 1, an audio encoder 100 includes a mask generator 102, a filter bank 104, a quantizer 106, and a bit stream generator 108. The audio encoder 100 executes an audio encoding process that generates encoded audio data 112 from input audio data 110. The encoded audio data 112 constitutes a compressed representation of the input audio

25 data 110.

The audio encoding process executed by the encoder 100 performs encoding steps based on MPEG-1-L2 processes described in the MPEG-1 standard. The time-domain input audio data 110 is converted into sub-bands by the filter bank 104, and the resulting

frequency-domain data is then quantized by the quantizer 106. The bitstream generator 108 then generates encoded audio data or bitstream 112 from the quantized data. The quantizer 106 performs bit allocation and quantization based upon masking data generated by the mask generator 102. The masking data is generated from the input audio data 110

5   on the basis of a psychoacoustic model of human hearing and aural perception. The psychoacoustic modelling takes into account the frequency-dependent thresholds of human hearing, and a psychoacoustic phenomenon referred to as masking, whereby a strong frequency component close to one or more weaker frequency components tends to mask, the weaker components, rendering them inaudible to a human listener. This makes it

10   possible to omit the weaker frequency components when encoding audio data, and thereby achieve a higher degree of compression, without adversely affecting the perceived quality of the encoded audio data 112. The masking data comprises a signal-to-mask ratio value for each frequency sub-band. These signal-to-mask ratio values represent the amount of signal masked by the human ear in each frequency sub-band. The quantizer 106 uses this

15   information to decide how best to use the available number of data bits to represent the input audio signal 110.

In known or prior art MPEG-1-L2 encoders, the generation of masking data has been found to be the most computationally intensive component of the encoding process,

20   representing up to 50% of the total processing resources. The MPEG-1 standard provides two example implementations of the psychoacoustic model: psychoacoustic model 1 (PAM1) is less complex and makes more compromises on quality than psychoacoustic model 2 (PAM2). PAM2 has better performance for lower bit rates. Nonetheless, quality tests indicate that PAM1 can achieve good quality encoding at high bit rates such as 256

25   and 384 kbps. However, PAM1 is implemented in floating point arithmetic and is not optimized for chip-based encoders. As described in G.A. Davidson *et. al.*, *Parametric Bit Allocation in a Perceptual Audio Coder*, 97th Convention of Audio Engineering Society, November 1994, it has been estimated that PAM1 demands more than 30 MIPS of computing power per channel.

30

Moreover, despite using the C double precision type throughout, the ISO implementation uses an extremely large number of arithmetic operations, each resulting in a loss of precision at each step of the psychoacoustic masking data generation process.

5 The psychoacoustic mask generation process 300 executed by the mask generator 102 provides an implementation of the psychoacoustic model that maintains quality whilst significantly reducing the computational requirements.

In order to most clearly describe the advantages of the psychoacoustic mask generation 10 process 300, the steps of the process are described below with reference to a prior art process 200 for generating psychoacoustic masking data, as described in the MPEG-1 standard.

In the described embodiment, the audio encoder is a standard digital signal processor 15 (DSP) such as a TMS320 series DSP manufactured by Texas Instruments. The audio encoding modules 102 to 108 of the encoder 100 are software modules stored in the firmware of the DSP-core. However, it will be apparent that at least part of the audio encoding modules 102 to 108 could alternatively be implemented as dedicated hardware components such as application-specific integrated circuits (ASICs).

20

As shown in Figures 2 and 3, both the psychoacoustic mask generation process 300 and the prior art process 200 for generating masking data begin by Hann windowing the 512-sample time-domain input audio data frame 110 at step 204. The Hann windowing effectively centers the 512 samples between the previous samples and the subsequent 25 samples, using a Hann window to provide a smooth taper. This reduces ringing edge artefacts that would otherwise be produced at step 206 when the time-domain audio data 110 is converted to the frequency domain using a 1024-point fast Fourier transform (FFT). At step 208, an array of 512 energy values for respective frequency sub-bands is then generated from the symmetric array of 1024 FFT output values, according to:

30 $$E(n) = |X(n)|^2 = X_R^2(n) + X_I^2(n) ,$$

where $X(n) = X_R(n) + iX_I(n)$ is the FFT output of the nth spectral line.

In this specification, a value or entity is described as logarithmic or as being in the logarithmic-domain if it has been generated as the result of evaluating a logarithmic function. When a logarithmic value or entity is exponentiated by the reverse operation, it is described as linear or as being in the linear-domain.

In the prior art process 200, the linear energy values $E(n)$ are then converted into logarithmic power spectral density (PSD) values $P(n)$ at step 210, according to $P(n) = 10\log_{10} E(n)$, and the linear energy values $E(n)$ are not used again. The PSD values are normalised to 96 dB at step 212.

Steps 210 and 212 are omitted from the mask generation process 300.

The next step in both processes is to generate sound pressure level (SPL) values for each sub-band. In the prior art process, an SPL value $L_{sb}(n)$ is generated for each sub-band n at step 214, according to:

$$L_{sb}(n) = MAX\left[X_{spl}(n), \quad 20*\log(scf_{max}(n)*32768) - 10\right]dB$$

and

$$X_{spl}(n) = 10*\log_{10}\left(\sum_k 10^{X(k)/10}\right)dB$$

where $scf_{max}(n)$ is the maximum of the three scale factors of sub-band n within an MPEG1 L2 audio frame comprising 1152 stereo samples, $X(k)$ is the PSD value of index k, and the summation over k is limited to values of k within sub-band n. The "-10 dB" term corrects for the difference between peak and RMS levels.

Significantly, the prior art generation of SPL values involves evaluating many exponentials and logarithms in order to convert logarithmic power values to linear energy values, sum

them, and then convert the summed linear energy values back to logarithmic power values. Each conversion between the logarithmic and linear domains is computationally expensive and degrades the precision of the result.

5   In the mask generation process $300, L_{sb}(n)$ is generated at step 302 using the same first formula for $L_{sb}(n)$, but with:

$$\dot{X}_{spl}(n) = 10 * \log_{10}\left(\sum_k X(k)\right) + 96 \, \text{dB}$$

where $X(k)$ is the linear energy value of index k. The "96 dB" term is used to
10   normalise $L_{sb}(n)$. It will be apparent that this improves upon the prior art by avoiding exponentiation. Moreover, the efficiency of generating the SPL values is significantly improved by approximating the logarithm by a second order Taylor expansion.

Specifically, representing the argument of the logarithm as Ipt, this is first normalised by
15   determining x such that:

$$Ipt = (1-x)2^m, \quad 0.5 < 1-x \le 1$$

Using a second order Taylor expansion,
$$\ln(1-x) \approx -x - x^2/2$$
20   the logarithm can be approximated as:
$$\log_{10}(Ipt) \approx [m * \ln(2) - (x + x^2/2)] * \log_{10}(e)$$
$$= [m * \ln(2) - (x + x * x * 0.5)] * \log_{10}(e)$$

--

25   Thus the logarithm is approximated by four multiplications and two additions, providing a significant improvement in computational efficiency.

The next step is to identify frequency components for masking. Because the tonality of a masking component affects the masking threshold, tonal and non-tonal (noise) masking components are determined separately.

5 First, local maxima are identified. A spectral line $X(k)$ is deemed to be a local maximum if

$$X(k) > X(k-1) \text{ and } X(k) \geq X(k+1)$$

In the prior art process 200, a local maximum $X(k)$ thus identified is selected as a
10 logarithmic tonal masking component at step 216 if:

$$X(k) - X(k+j) \geq 7 \, \text{dB}$$

where j is a searching range that varies with k. If $X(k)$ is found to be a tonal component, then its value is replaced by:

15 $$X_{tonal}(k) = 10 \log_{10} \left(10^{X(k-1)/10} + 10^{X(k)/10} + 10^{X(k+1)/10} \right)$$

All spectral lines within the examined frequency range are then set to $-\infty$ dB.

In the mask generation process 300, a local maximum $X(k)$ is selected as a linear tonal
20 masking component at step 304 if:

$$X(k) * 10^{-0.7} \geq X(k+j)$$

If $X(k)$ is found to be a tonal component, then its value is replaced by:

$$X_{tonal}(k) = X(k-1) + X(k) + X(k+1)$$

25 All spectral lines within the examined frequency range are then set to 0.

The next step in either process is to identify and determine the intensity of non-tonal masking components within the bandwidth of critical sub-bands. For a given frequency,

the smallest band of frequencies around that frequency which activate the same part of the basilar membrane of the human ear is referred to as a critical band. The critical bandwidth represents the ear's resolving power for simultaneous tones. The bandwidth of a sub-band varies with the center frequency of the specific critical band. As described in the MPEG-1 standard, 26 critical bands are used for a 48 kHz sampling rate. The non-tonal (noise) components are identified from the spectral lines remaining after the tonal components are removed as described above.

At step 218 of the prior art process 200, the logarithmic powers of the remaining spectral lines within each critical band are converted to linear energy values, summed and then converted back into a logarithmic power value to provide the SPL of the new non-tonal component $X_{noise}(k)$ corresponding to that critical band. The number k is the index number of the spectral line nearest to the geometric mean of the critical band.

In the mask generation process 300, the energy of the remaining spectral lines within each critical band are summed at step 306 to provide the new non-tonal component $X_{noise}(k)$ corresponding to that critical band:

$$X_{noise}(k) = \sum_k X(k)$$

for k in sub-band n. Only addition is used, and no exponential or logarithmic evaluations are required, providing a significant improvement in efficiency.

The next step is to decimate the tonal and non-tonal masking components. Decimation is a procedure that is used to reduce the number of masking components that are used to generate the global masking threshold.

In the prior art process 200, logarithmic tonal components $X_{tonal}(k)$ and non-tonal components $X_{noise}(k)$ are selected at step 220 for subsequent use in generating the masking threshold only if:

$$X_{tonal}(k) \geq LT_q(k) \quad \text{or} \quad X_{noise}(k) \geq LT_q(k)$$

respectively, where $LT_q(k)$ is the absolute threshold (or threshold in quiet) at the frequency of index k; threshold in quiet values in the logarithmic domain are provided in the MPEG-1 standard.

5   Decimation is performed on two or more tonal components that are within a distance of less than 0.5 Bark, where the Bark scale is a frequency scale on which the frequency resolution of the ear is approximately constant, as described in E. Zwicker, *Subdivision of the Audible Frequency Range into Critical Bands*, J. Acoustical Society of America, vol. 33, p. 248, February 1961. The tonal component with the highest power is kept while the

10   smaller component(s) are removed from the list of selected tonal components. For this operation, a sliding window in the critical band domain is used with a width of 0.5 Bark.

In the mask generation process 300, linear components are selected at step 308 only if:

15
$$X_{tonal}(k) \geq LT_q E(k) \quad \text{or} \quad X_{noise}(k) \geq LT_q E(k)$$

where $LT_q E(k)$ are taken from a linear-domain absolute threshold table pre-generated from the logarithmic domain absolute threshold table $LT_q(k)$ according to:

$$LT_q E(k) = 10^{\log_{10}\lfloor LT_q(k)-96\rfloor / 10}$$

20
where the "-96" term represents denormalization.

After denormalization, the spectral data in the linear energy domain are converted into the logarithmic power domain at step 310. In contrast to step 206 of the prior art process, the

25   evaluation of logarithms is performed using the efficient second-order approximation method described above. This conversion is followed by normalization to the reference level of 96 dB at step 212.

Having selected and decimated masking components, the next step is to generate

30   individual masking thresholds. Of the original 512 spectral data values, indexed by k, only

a subset, indexed by i, is subsequently used to generate the global masking threshold, and this step determines that subset by subsampling, as described in the MPEG-1 standard.

The number of lines n in the subsampled frequency domain depends on the sampling rate.
5 For a sampling rate of 48 kHz, $n = 126$. Every tonal and non-tonal component is assigned an index i that most closely corresponds to the frequency of the corresponding spectral line in the original (*i.e.*, before sub-sampling) spectral data.

The individual masking thresholds of both tonal and non-tonal components, LT_{tonal} and
10 LT_{noise}, are then given by the following expressions:

$$LT_{tonal}[z(j), z(i)] = X_{tonal}[z(j)] + av_{tonal}[z(j)] + vf[z(j), z(i)] \quad dB$$

$$LT_{noise}[z(j), z(i)] = X_{noise}[z(j)] + av_{noise}[z(j)] + vf[z(j), z(i)] \quad dB$$

15 where i is the index corresponding to a spectral line, at which the masking threshold is generated and j is that of a masking component; $z(i)$ is the Bark scale value of the ith spectral line while $z(j)$ is that of the jth line; and terms of the form $X[z(j)]$ are the SPLs of the (tonal or non-tonal) masking component. The term av, referred to as the masking index, is given by:
20

$$av_{tonal} = -1.525 - 0.275 * z(j) - 4.5 \quad dB$$

$$av_{noise} = -1.525 - 0.175 * z(j) - 0.5 \quad dB$$

vf is a masking function of the masking component and is characterized by different lower
25 and upper slopes, depending on the distance in Bark scale dz, $dz = z(i) - z(j)$

In the prior art process 200, individual masking thresholds are generated at step 222 using a masking function vf given by:

30 $\qquad vf = 17 * (dz + 1) - 0.4 * X[z(j)] - 6 \, dB, \quad \text{for} - 3 \le dz < -1 \text{ Bark}$

$$vf = \{0.4 * X[z(j)] + 6\} * dz \text{ dB}, \quad \text{for } -1 \le dz < 0 \text{ Bark}$$

$$vf = -17 * dz \text{ dB}, \quad \text{for } 0 \le dz < 1 \text{ Bark}$$

5 $$vf = -17 * dz + 0.15 * X[z(j)] * (dz - 1) \text{ dB}, \quad \text{for } 1 \le dz < 8 \text{ Bark}$$

where $X[z(j)]$ is the SPL of the masking component with index j. No masking threshold is generated if $dz < -3$ Bark, or $dz > 8$ Bark.

10   The evaluation of the masking function vf is the most computationally intensive part of this step of the prior art process. The masking function can be categorized into two types: downward masking (when $dz < 0$) and upward masking (when $dz \ge 0$). As described in Davis Pan, *A Tutorial on MPEG/Audio Compression*, IEEE Journal on Multimedia, 1995, downward masking is considerably less significant than upward masking. Consequently,
15   only upward masking is used in the mask generation process 300. Moreover, further analysis shows that the second term in the masking function for $1 \le dz < 8$ Bark is typically approximately one tenth of the first term, $-17*dz$. Consequently, the second term can be safely discarded.

20   Accordingly, the mask generation process 300 generates individual masking thresholds at step 312 using a single expression for the masking function vf, as follows:

$$vf = -17 * dz, \quad 0 \le dz < 8$$

This greatly reduces the computational load while maintaining good quality encoding. The masking index av is not modified from that used in the prior art process, because it makes
25   a significant contribution to the individual masking threshold LT and is not computationally demanding.

After the individual masking thresholds have been generated, a global masking threshold is generated.
30

In the prior art process 200, the global masking threshold $LT_g(i)$ at the ith frequency sample is generated at step 224 by summing the powers corresponding to the individual masking thresholds and the threshold in quiet, according to:

5

$$LT_g(i) = 10\log_{10}\left[10^{LT_q(i)/10} + \sum_{j=1}^{m}10^{LT_{tonal}[z(j),z(i)]}/10 + \sum_{j=1}^{n}10^{LT_{noise}[z(j),z(i)]}/10\right]$$

where m is the total number of tonal masking components, and n is the total number of non-tonal masking components. The threshold in quiet LT_q is offset by -12 dB for bit rates ≥ 96 kbps per channel.

10

It will be apparent that this step is computationally demanding due to the number of exponentials and logarithms that are evaluated.

In the mask generation process 300, these evaluations are avoided and smaller terms are

15   not used. The global masking threshold $LT_g(i)$ at the ith frequency sample is generated at step 314 by comparing the powers corresponding to the individual masking thresholds and the threshold in quiet, as follows:

$$LT_g(i) = \max\left[LT_q(i) + \max_{j=1}^{m}\{LT_{tonal}[z(j),z(i)]\} + \max_{j=1}^{n}\{LT_{noise}[z(j),z(i)]\}\right]$$

20

The largest tonal masking components and of non-tonal masking components are identified. They are then compared with $LT_q(i)$. The maximum of these three values is selected as the global masking threshold at the ith frequency sample. This reduces computational demands at the expense of occasional over allocation. As above, the

25   threshold in quiet LT_q is offset by -12dB for bit rates ≥ 96 kbps per channel.

Finally, signal-to-mask ratio values are generated at step 226 of both processes. First, the minimum masking level $LT_{min}(n)$ in sub-band n is determined by the following expression:

$$LT_{\min}(n) = Min[LT_g(i)] \quad \text{dB;} \quad \text{for } f(i) \text{ in subband } n,$$

where $f(i)$ is the ith frequency line within sub-band n. A minimum masking threshold
$LT_{\min}(n)$ is determined for every sub-band. The signal-to-mask ratio for every sub-band n
is then generated by subtracting the minimum masking threshold of that sub-band from the
corresponding SPL value:

$$SMR_{sb}(n) = L_{sb}(n) - LT_{\min}(n)$$

The mask generator 102 sends the signal-to-mask ratio data $SMR_{sb}(n)$ for each sub-band n
to the quantizer 104, which uses it to determine how to most effectively allocate the
available data bits and quantize the spectral data, as described in the MPEG-1 standard.

Many modifications will be apparent to those skilled in the art without departing from the
scope of the present invention as herein described with reference to the accompanying
drawings.

CLAIMS:

1. A mask generation process for use in encoding audio data, including:

   generating linear masking components from said audio data;

5   generating logarithmic masking components from said linear masking components; and

   generating a global masking threshold from the logarithmic masking components.

10 2. A mask generation process as claimed in claim 1, wherein said step of generating linear masking components includes:

   generating linear components in a frequency domain from said audio data;

   selecting a first subset of said linear components as linear tonal components; and

15   selecting a second subset of said linear components as linear non-tonal components.

3. A mask generation process as claimed in claim 2, including generating sound pressure levels from said linear components using a second-order Taylor expansion of a

20 logarithmic function.

4. A mask generation process as claimed in claim 1, wherein said logarithmic masking components are generated using a second-order Taylor expansion of a logarithmic function.

25

5. A mask generation process as claimed in claim 3 or 4, including generating a normalised value corresponding to an argument of said logarithmic function, and using said normalised value in said Taylor expansion.

30 6. A mask generation process as claimed in claim 5, including:

   generating said normalised value x for said argument Ipt, according to:

$$Ipt = (1-x)2^m, \quad 0.5 < 1 - x \leq 1$$

and using a second order Taylor expansion of the form

$$\ln(1-x) \approx -x - x^2/2$$

to approximate said logarithmic function as:

$$\log_{10}(Ipt) \approx [m * \ln(2) - (x + x^2/2)] * \log_{10}(e)$$

7. A mask generation process as claimed in claim 2, wherein said step of generating a global masking threshold includes:

decimating said linear tonal components and said linear non-tonal components; and

generating masking thresholds from the decimated linear tonal components and the decimated linear non-tonal components.

8. A mask generation process as claimed in claim 1, including generating masking thresholds from said logarithmic masking components using a masking function of the form:

$$vf = -17 * dz, \quad 0 \leq dz < 8$$

9. A mask generation process as claimed in claim 7, wherein said step of generating a global masking threshold includes determining maximum components of said masking thresholds and predetermined threshold values.

10. A mask generation process as claimed in claim 9, wherein said global masking threshold is generated according to:

$$LT_g(i) = \max\left[LT_q(i) + \max_{j=1}^m \{LT_{tonal}[z(j), z(i)]\} + \max_{j=1}^n \{LT_{noise}[z(j), z(i)]\}\right]$$

where i and j are indices of logarithmic power components, $z(i)$ is a Bark scale value for logarithmic power component i, $LT_{tonal}[z(j), z(i)]$ is a tonal masking threshold for logarithmic power components i and j, $LT_{noise}[z(j), z(i)]$ is a non-tonal masking threshold for logarithmic power components i and j, m is the number of tonal

logarithmic power components, and n is the number of non-tonal logarithmic power components.

11. A mask generation process for use in encoding audio data, including:

      generating respective masking thresholds from logarithmic masking components using a masking function of the form:

$$vf = -17 * dz, \quad 0 \leq dz < 8$$

12. A mask generation process for use in encoding audio data, including:

      generating a global masking threshold from logarithmic masking components according to:

$$LT_g(i) = \max\left[LT_q(i) + \max_{j=1}^{m}\{LT_{tonal}[z(j), z(i)]\} + \max_{j=1}^{n}\{LT_{noise}[z(j), z(i)]\}\right]$$

where i and j are indices of spectral audio data, $z(i)$ is a Bark scale value for spectral line i, $LT_{tonal}[z(j), z(i)]$ is a tonal masking threshold for lines i and j, $LT_{noise}[z(j), z(i)]$ is a non-tonal masking threshold for lines i and j, m is the number of tonal spectral lines, and n is the number of non-tonal spectral lines.

13. A mask generation process as claimed in any one of claims 1 to 13, wherein said linear masking components include linear energy components, and said logarithmic masking components include logarithmic power components.

14. A mask generation process as claimed in any one of claims 1 to 12, wherein said process is an MPEG-1 layer 2 audio encoding process.

15. A mask generator having components for executing the steps of any one of claims 1 to 14.

16. An audio encoder having components for executing the steps of any one of claims 1 to 14.

17. A computer readable storage medium having stored thereon program code for
    executing the steps of any one of claims 1 to 14.

18. A mask generator for an audio encoder, said mask generator adapted to generate linear
    masking components from input audio data, logarithmic masking components from
    said linear masking components; and a global masking threshold from the logarithmic
    masking components.

19. A psychoacoustic masking process for use in an audio encoder, including:
    generating energy values from Fourier transformed audio data;
    determining sound pressure level values from said energy values;
    selecting tonal and non-tonal masking components on the basis of said energy values;
    generating power values from said energy values;
    generating masking thresholds on the basis of said masking components and said
    power values; and
    generating signal to mask ratios for a quantizier on the basis of said sound pressure
    level values and said masking thresholds.

20. An MPEG-1-L2 encoder adapted to execute the masking process of claim 19.

162162

G00002

ABSTRACT

5   **DEVICE AND PROCESS FOR USE IN ENCODING AUDIO DATA**

A mask generation process for use in encoding audio data, including generating linear masking components from the audio data, generating logarithmic masking components from the linear masking components, and generating a global masking threshold from the

10   logarithmic masking components. The process is a psychoacoustic masking process for use in an MPEG-1-L2 encoder, and includes generating energy values from a Fourier transform of the audio data, determining sound pressure level values from the energy values, selecting tonal and non-tonal masking components on the basis of the energy values, generating power values from the energy values, generating masking thresholds on

15   the basis of the masking components and the power values, and generating signal to mask ratios for a quantizier on the basis of the sound pressure level values and the masking thresholds.
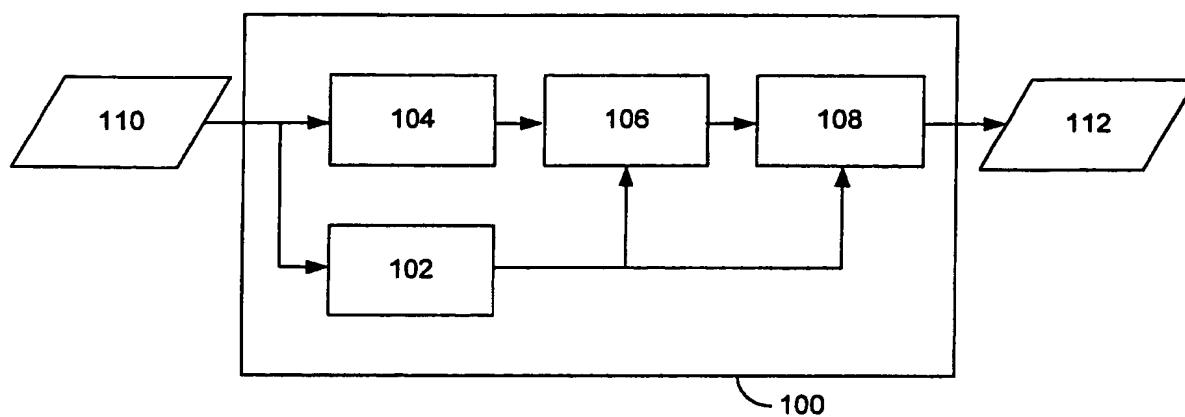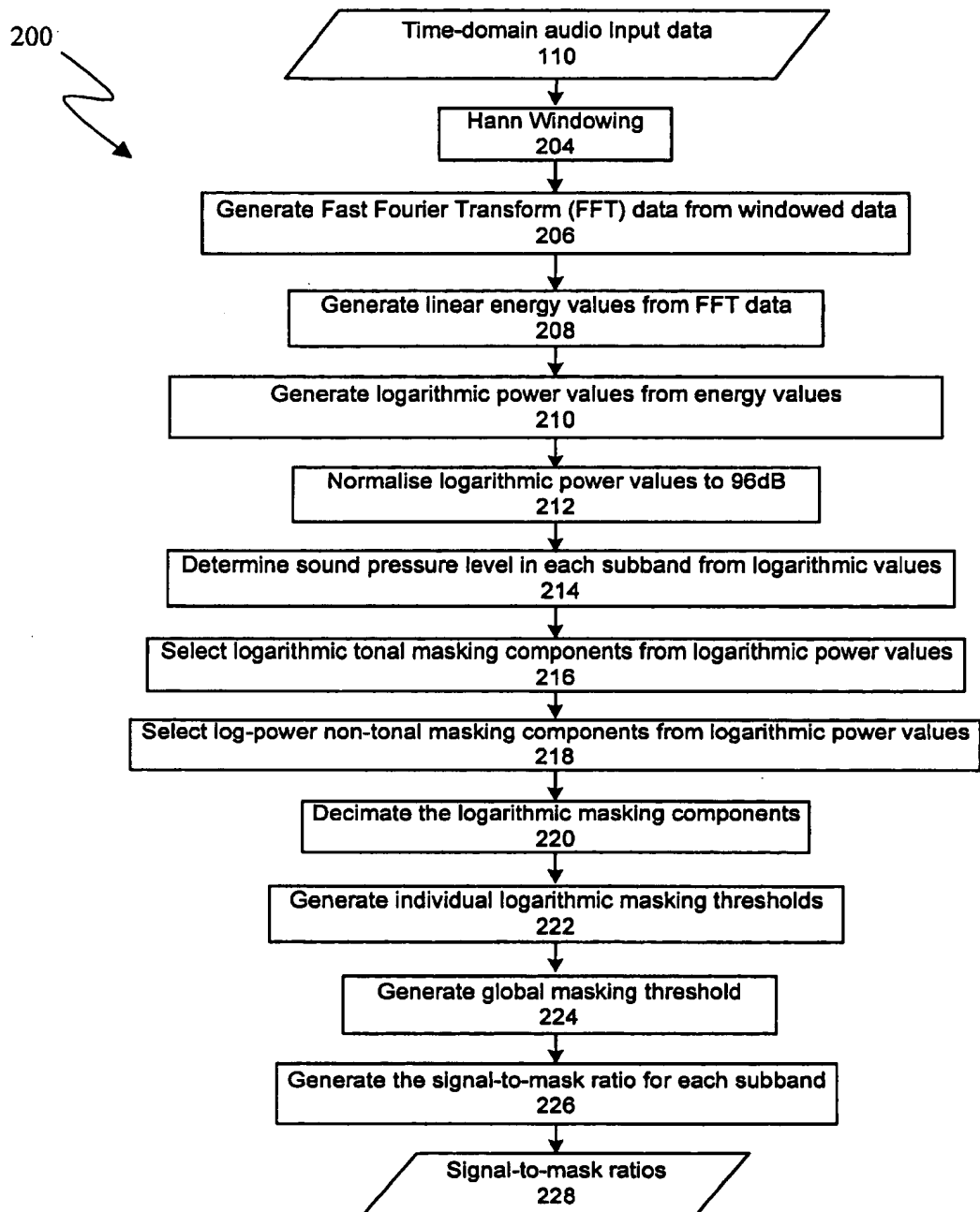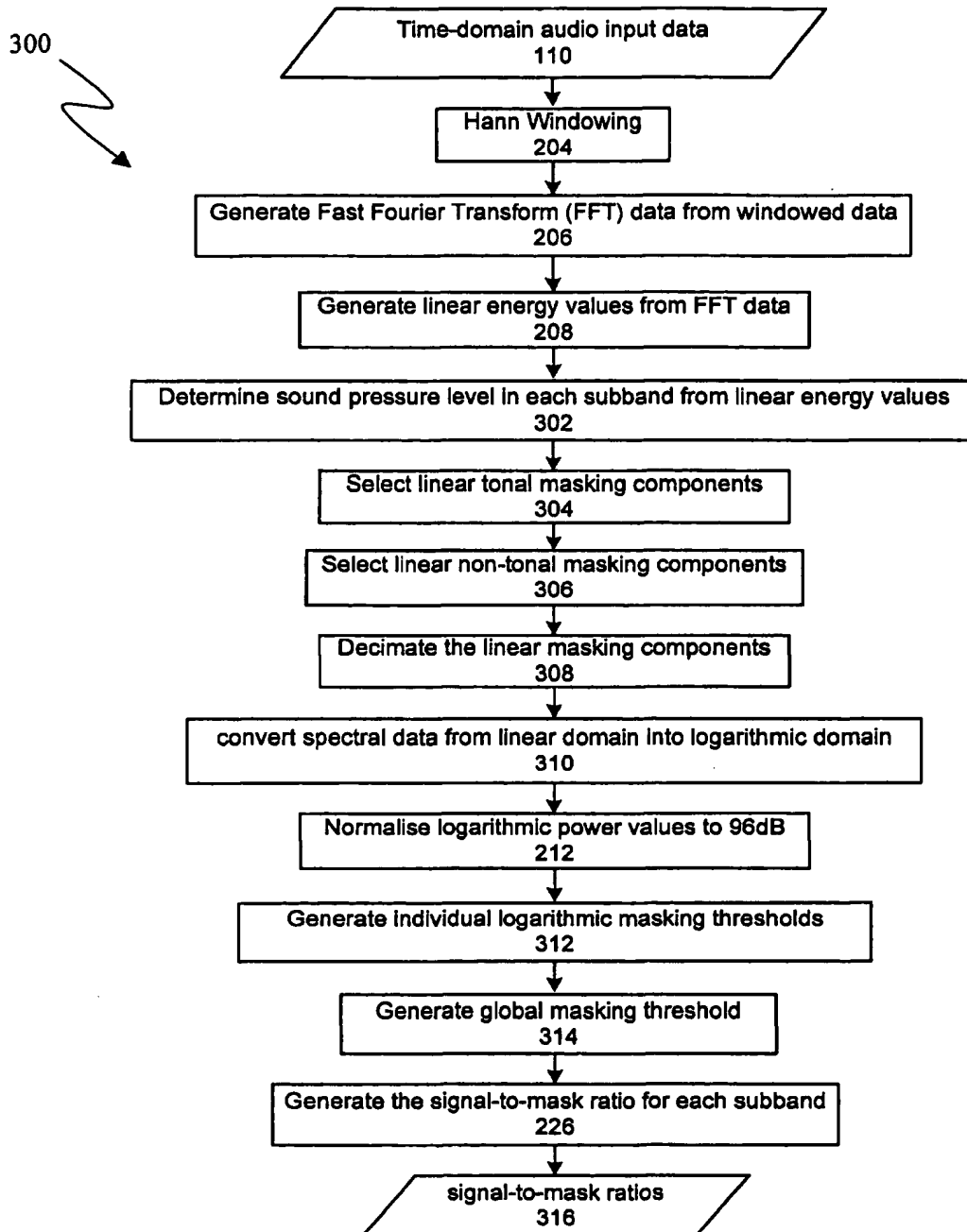
Figure 1

Figure 2

Figure 3